Toxicology in Vitro xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Toxicology in Vitro

journal homepage: www.elsevier.com/locate/toxinvit

A comprehensive statistical classifier of *foci* in the cell transformation assay for carcinogenicity testing

Giulia Callegaro[a], Kasja Malkoc[b], Raffaella Corvi[c], Chiara Urani[a,*], Federico M. Stefanini[b,**]

[a] Department of Earth and Environmental Sciences, University of Milan Bicocca, Piazza della Scienza, 1, 20126 Milan, Italy
[b] Department of Statistics, Computer Science, Applications, University of Florence, Viale Morgagni 59, 50134 Florence, Italy
[c] European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM), Chemical Safety and Alternative Methods Unit, Directorate F, European Commission Joint Research Centre, TP 126, Via E. Fermi 2749, I-21027 Ispra, VA, Italy

ARTICLE INFO

Keywords:
BALB/c 3T3 Cell Transformation Assay
In vitro carcinogenicity testing
Foci classification
Statistical image descriptors
Bayesian Trees

ABSTRACT

The identification of the carcinogenic risk of chemicals is currently mainly based on animal studies. The *in vitro* Cell Transformation Assays (CTAs) are a promising alternative to be considered in an integrated approach. CTAs measure the induction of *foci* of transformed cells. CTAs model key stages of the *in vivo* neoplastic process and are able to detect both genotoxic and some non-genotoxic compounds, being the only *in vitro* method able to deal with the latter. Despite their favorable features, CTAs can be further improved, especially reducing the possible subjectivity arising from the last phase of the protocol, namely visual scoring of *foci* using coded morphological features. By taking advantage of digital image analysis, the aim of our work is to translate morphological features into statistical descriptors of *foci* images, and to use them to mimic the classification performances of the visual scorer to discriminate between transformed and non-transformed *foci*. Here we present a classifier based on five descriptors trained on a dataset of 1364 *foci*, obtained with different compounds and concentrations. Our classifier showed accuracy, sensitivity and specificity equal to 0.77 and an area under the curve (AUC) of 0.84. The presented classifier outperforms a previously published model.

1. Introduction

Cancer is among the leading causes of morbidity and mortality worldwide, and it is expected that morbidity will further increase within the next two decades (Stewart and Wild, 2014). Cancer development is associated and influenced by various factors, including genetic predisposition and exposure to chemicals. For the latter reason the evaluation of carcinogenicity, defined as the ability for a chemical substance, or a mixture of chemical substances, to induce cancer or increase its incidence, has played a major role in public health and risk assessment research throughout the last decades (IARC, 2004; U.S. EPA, 2003).

Carcinogenicity evaluation of chemicals usually requires a battery of *in vitro* and *in vivo* genotoxicity tests that can be followed by the lifetime cancer rodent bioassay (OECD, 2009a, 2009b). Acknowledging the several limitations and concerns of the *in vivo* bioassays (Paules et al., 2011; Knight et al., 2006a, 2006b), Cell Transformation Assays (CTAs) represent a promising *in vitro* alternative for carcinogenicity testing to be considered as a component of an Integrated Approach (Jacobs et al., 2016; Benigni, 2014; Jaworska and Hoffmann, 2010).

The multistep process of *in vitro* transformation can model the progressive nature of the *in vivo* carcinogenesis through similar cellular and molecular events (Barrett and Ts'o, 1978; Maurici et al., 2005) and the *in vitro* transformed cells were proved to be tumorigenic when inoculated into mice (Kakunaga, 1973; Keshava, 2000; Sasaki et al., 2012a, 2012b). Finally, CTAs are faster and cost-efficient in comparison to the rodent bioassay, their results can be informative of cellular and multicellular interactions, and to date represent the only standardized *in vitro* methods to detect genotoxic and some non-genotoxic carcinogens (OECD, 2007; Vanparys et al., 2012; Jacobs et al., 2016).

CTAs are currently used by industry and academia for screening purposes, they represent a tool for studying cancer mechanisms and

Abbreviations: AUC, Area Under the Curve; BD, Boundary Index; CTA, *in vitro* Cell Transformation Assay; ED, Equivalent Diameter; EFP, perimeter of circle with equivalent area; FP, *focus* perimeter; HD, Heterogeneity Detector; ITS, Integrated Testing Strategies; LSD, Line Segment Detector; MD, Median; ROI, Region of Interest; ROC, Receiver Operating Characteristic curve; SD, Multicellular-Spindle Detector

* Correspondence to: C. Urani, Department of Earth and Environmental Sciences, University of Milan Bicocca, Piazza della Scienza, 1-20126 Milan, Italy.
** Correspondence to: F.M. Stefanini, Department of Statistics, Computer Science, Applications "Giuseppe Parenti" (DiSIA), University of Florence, Viale Morgagni, 59, 50134 Firenze, Italy.
E-mail addresses: chiara.urani@unimib.it (C. Urani), stefanini@disia.unifi.it (F.M. Stefanini).

http://dx.doi.org/10.1016/j.tiv.2017.04.030
Received 21 December 2016; Received in revised form 13 April 2017; Accepted 28 April 2017

therapies (Forcella et al., 2016; Poburski and Thierbach, 2016), and further applications are the investigation of the carcinogenicity potential of environmental mixtures, pesticides and even radiations (Perocco et al., 1993; Balcer-Kubiczek et al., 1996; Mascolo et al., 2010; Rodríguez-Sastre et al., 2014; Colacci et al., 2014).

Nonetheless, further improvements may promote a wider implementation of CTAs for regulatory purposes. In particular, the mechanisms underlying *in vitro* morphological transformation necessitate in-depth analysis, and the possible subjectivity arising from the final phase of the protocol (the visual scoring of *foci* of transformed cells) needs to be completely dealt with (EURL ECVAM, 2012; Combes, 2012).

By taking advantage of digital image analysis, in recent years we have worked on the translation of morphological features used for visual scoring (Urani et al., 2013) into statistical descriptors of *foci* images, with the final aim of mimicking the performances of a visual scorer with an unbiased and quantitative approach. We previously applied a set of statistical descriptors of *foci* morphologies to build a classifier for *foci* obtained with exposure to methylcholanthrene (Callegaro et al., 2015). Another innovative aspect and application of our approach was the identification of a relationship between morphologies of transformed colonies (*foci*) and the carcinogens concentration in a quantitative way (Callegaro et al., 2016).

Here we present a comprehensive automated and objective classifier, trained with *foci* images obtained testing different compounds. In addition to the previously developed descriptors (Urani et al., 2013), we translated two more morphological features of *foci* (spindle-shape of transformed cells and texture heterogeneity of *foci*) into statistical descriptors. Texture heterogeneity is a key feature that could help to the identification of mixed and intermediate *foci*, that still represents an open problem in visual classification (Landolph, 1985). We combined all of them into a classifier supporting visual scoring of malignant *foci* in the BALB/c 3T3 CTA in current applications.

2. Materials and methods

2.1. Cell Transformation Assays: from plates to images

The plates used for image analysis of transformed *foci* were obtained from CTAs performed by EURL ECVAM within the prevalidation study of BALB/c 3T3 CTA (Corvi et al., 2012; Sasaki et al., 2012a). At the end of the experimental phase, *foci* being more than about 2 mm in diameter were evaluated by stereomicroscopy according to predefined morphological features, as detailed in the recommended protocol and photo catalogue (Sasaki et al., 2012b; Tanaka et al., 2012). According to the protocol, only Type III *foci* were considered as fully transformed thus scored as positive. Type III *foci* are characterized by the following morphological features: *foci* cells are deeply basophilic stained, spindle-shaped and morphologically different from the background monolayer of non-transformed cells. Type III *foci* show dense multilayering (piling up), random orientation and invasive growth of cells at the edge of *foci* (criss-cross pattern). *Foci* images were acquired by adopting the procedure previously developed by Callegaro et al. (2015) and briefly described below.

Acquisition was performed under a stereomicroscope (Zeiss, Stemi SV6) equipped with 6.3 × lens (Carl Zeiss, Arese, Italy) and a digital camera (AxioCam Mrc5, 36 bit). Images were saved in TIFF-48bit RGB format and had a size of 2572 × 1928 pixels, where 1 pixel is equivalent to a real size of 6.7842×10^{-3} mm (1 cm = 1474 pixels). Both fully transformed Type III *foci* and non-Type III *foci* (Type I and II) were considered for the acquisition, as well as *foci* obtained from various coded and uncoded compounds and concentrations, as detailed in Table 1. An original segmentation algorithm coded in Python (Python Software Foundation, https://www.python.org/) was applied to each *focus* image in order to isolate the *focus* region from the surrounding monolayer (Callegaro et al., 2015). The algorithm acquires a RGB *focus* image, converts it into the HSV color space and separates the Region of Interest (ROI) corresponding to each *focus* from the surrounding monolayer by setting appropriate intensity thresholds.

Table 1
Dataset composition. Dataset composition is detailed for compound and number of Type III *foci* and non-Type III *foci*. For each compound, the number of concentrations tested is shown. Compounds tested were: dimethyl sulfoxide 0.5% (DMSO, #67–68-5), 3-Methylcholanthrene (MCA, #56–49-5), from 0.01 to 10 μg/ml, 2-Acetylaminofluorene (A, #53–96-3), from 0.05 to 35 μg/ml, Benzo[*a*]Pyrene (B, #50–32-8), from 0.0005 to 15 μg/ml, Anthracene (HB, #120–12-7), from 0.1 to 40 μg/ml, o-toluidine (TB, #636–21-5), from 20 to 1750 μg/ml, and $NiCl_2$ (#50–32-8), from 50 to 400 μM.

	Concentrations tested	Number of Type III *foci*	Number of non-Type III *foci*
Control	–	11	9
DMSO	–	16	11
MCA	7	487	330
A	8	150	61
B	9	224	174
HB	6	47	15
TB	8	84	23
$NiCl_2$	8	369	6
Total	–	1416	630

At the end of the process of segmentation, the final dataset comprised a total of 2046 *foci* images, including both Type III and non-Type III *foci*, in the proportion detailed in Table 1.

2.2. Quantifying morphological features: Equivalent Diameter, Median and Boundary Index descriptors

From each *focus* ROI the set of previously developed descriptors was calculated (Urani et al., 2013). These three descriptors are related to three of the morphological features used for visual scoring in BALB/c 3T3 CTA (Sasaki et al., 2012a). Specifically, the Equivalent Diameter (ED) is the diameter of the circle having the same area of the *focus*, then calculated as: $= 2\sqrt{\frac{ared}{\pi}}$, where π is the trigonometric constant. ED measures *foci* dimensions, taking into account their shape polymorphism. The median of the gray-scale image pixels distribution (MD) was selected as a proxy of *focus* ability to grow into multilayers, as image gray levels (or saturation) depend on the amount of light passing through *focus* region. The more a *focus* is multilayered, the darker will be its grayscale image. Finally, the descriptor called Boundary Index (BD) was developed in order to capture a trait of *foci* invasiveness. Invasiveness often implies boundary heterogeneity of *foci*, thus BD compares the actual *focus* perimeter (FP) and the perimeter of the circle having the same *focus* area (EFP): $BD = (\frac{FP}{EFP}) - 1$.

2.3. Multicellular-spindle Detector (SD): a procedural workflow

BALB/c 3T3 cells originate from mesenchimal fibroblastic cells; at the early stage of culture they are spindle-shaped, while they exhibit an epithelial-like morphology (cobble-stones) when they are at the contact-inhibited confluent stage. In the CTA, malignantly transformed cells change morphology from epithelial-like (as the cells growing in the background monolayer) to rather spindle-shaped, probably because of rearranged structures and functions in the transformation process, which lead to *focus* formation. The “spindle-shape” is among the coded morphological features for *foci* visual scoring (Sasaki et al., 2012b). However, at the recommended magnification (50 ×, Sasaki et al., 2012b), single cell shapes are not always observable. Nonetheless, spindle-shaped cells tend to pack into macro-segments that are clearly visible and detectable at the magnification suggested. See as examples the aggregate regions of pictures 41, 42 and 47 provided in the photo catalogue for the classification of *foci* (Sasaki et al., 2012b).

We therefore applied a Line Segment Detector, LSD (Grompone von Gioi et al., 2012) to each gray-scale *focus* image (step 2a, Fig. 1). LSD is a linear-time Line Segment Detector aiming at detecting locally straight